Identification of novel gene expression signature in lung adenocarcinoma by using next-generation sequencing data and bioinformatics analysis

Ya-Ling Hsu1,*, Jen-Yu Hung2,3,*, Yen-Lung Lee4, Feng-Wei Chen5, Kuo-Feng Chang6, Wei-An Chang3,5, Ying-Ming Tsai1,3, Inn-Wen Chong3,7 and Po-Lin Kuo5,8,9

1 Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
2 School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
3 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
4 Division of Thoracic Surgery, Department of Surgery, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan
5 Graduate Institute of Clinical Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
6 Welgene Biotech. Inc, Taipei, Taiwan
7 Department of Respiratory Therapy, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
8 Center for Biomarkers and Biotech Drugs, Kaohsiung Medical University, Kaohsiung, Taiwan
9 Institute of Medical Science and Technology, National Sun Yat-Sen University, Kaohsiung, Taiwan

* These authors have contributed equally to this work

Correspondence to: Inn-Wen Chong, email: chong@kmu.edu.tw
Po-Lin Kuo, email: kuopolin@seed.net.tw

Keywords: next-generation sequencing, bioinformatics, microRNA, messenger RNA, lung adenocarcinoma

Received: July 31, 2017 Accepted: August 28, 2017 Published: September 18, 2017

ABSTRACT

Lung adenocarcinoma is one of the leading causes of cancer-related death worldwide. We showed transcriptomic profiles in three pairs of tumors and adjacent non-tumor lung tissues using next-generation sequencing (NGS) to screen protein-coding RNAs and microRNAs. Combined with meta-analysis from the Oncomine and Gene Expression Omnibus (GEO) databases, we identified a representative genetic expression signature in lung adenocarcinoma. There were 9 upregulated genes, and 8 downregulated genes in lung adenocarcinoma. The analysis of the effects from each gene expression on survival outcome indicated that 6 genes (AGR2, SPDEF, CDKN2A, CLDN3, SFN, and PHLDA2) play oncogenic roles, and 7 genes (PDK4, FMO2, CPED1, GNG11, IL33, BTNL9, and FABP4) act as tumor suppressors in lung adenocarcinoma. In addition, we also identified putative genetic interactions, in which there were 5 upregulated microRNAs with specific targets - hsa-miR-183-5p-BTNL9, hsa-miR-33b-5p-CPED1, hsa-miR-429-CPED1, hsa-miR-182-5p-FMO2, and hsa-miR-130b-5p-IL33. These 5 microRNAs have been shown to be associated with tumorigenesis in lung cancer. Our findings suggest that these genetic interactions play important roles in the progression of lung adenocarcinoma. We propose that this molecular change of genetic expression may represent a novel signature in lung adenocarcinoma, which may be developed for diagnostic and therapeutic strategies in the future.

INTRODUCTION

Lung cancer is one of the leading causes of cancer-related death worldwide [1]. The development and progression of lung cancer has been widely studied. Briefly, the genetic alterations or mutations occurred in a single cell, leading to cellular transformation and thus expansion into a malignant tumor [2]. Non-small cell
lung carcinoma (NSCLC) accounts for about 80–85% of all lung cancers [3, 4], of which lung adenocarcinoma (40%) is the most common subtype of NSCLC [5]. Surgery to remove cancer, with or without radiotherapy/chemotherapy, is the standard approach for early stage lung cancer. However, the recurrence of distant metastasis [6, 7] or resistance to therapy [8, 9] often occurs, and such phenomenon is associated with critical genetic alterations involved in various biological mechanisms.

The genetic alterations related to cellular transformation are involved in various biological processes, including transcription [10], DNA repair [11], cell cycle progression [12], apoptosis [13], migration ability [14], and metabolism [15, 16]. In lung cancer, many genes have been identified as oncogenes or tumor suppressor genes, which modulate various of molecular functions involved in tumor development and progression [17, 18]. Recently, small RNAs have been found to play important roles in lung cancer. MicroRNAs are a group of small non-coding RNAs containing 20–26 nucleotides, which can regulate gene expression via binding to the 3’ untranslated region (3’UTR) of specific messenger RNAs (mRNAs). This interaction can cause mRNAs’ degradation or translation inhibition [19]. The signaling pathways associated with microRNAs targeting oncogenes or tumor suppressor genes have been reported to be involved in lung cancer progression [20]. Alterations of many microRNAs in chromosome regions associated with various cancers have also been implicated [21].

Next-generation sequencing (NGS) is a powerful method to screen the entire transcriptomic profile, including messenger RNAs or small RNAs [22]. In this study, we attempted to identify the differentially expressed genes and genetic interactions of target gene-microRNA in lung adenocarcinoma combined with systematic analysis, by using bioinformatics tools, including the Oncomine database [23], Gene Expression Omnibus (GEO) [24], PrognoScan [25], Kaplan–Meier plotter [26], SurvExpress [27], and miRmap databases [28]. We sought to identify novel gene expression signature and/or genetic interactions in lung adenocarcinoma via systematic bioinformatics analysis. Hopefully, the approach and findings from this study will provide new perspectives on the development of diagnostic and therapeutic strategies for lung adenocarcinoma.

RESULTS

Identification of differentially expressed genes as a molecular signature in lung adenocarcinoma

To investigate genetic expression changes in lung adenocarcinoma, we analyzed the transcriptomic profiles in 3 pairs of human specimens from lung adenocarcinoma and its adjacent normal lung tissue using next-generation sequencing (Figure 1). We focused on protein-coding RNAs and Venn diagram analysis which showed that 9 genes were upregulated (Figure 1A), whereas 8 genes were downregulated (Figure 1B) in lung adenocarcinoma tissue compared to adjacent normal lung tissue. The analysis criteria were fold change > 2 and fragments per kilobase million (FPKM) > 0.3. The hierarchical color clustering showed the expression pattern of each gene with z-scores (log2) in these 3 pairs of specimens (Figure 1C). The list of 17 differentially expressed genes with FPKM is shown in Table 1. To investigate whether this genetic expression pattern could represent a molecular signature in lung adenocarcinoma, we investigated these genes in the Oncomine database, which contains different data sets of specimens from lung adenocarcinoma and normal lung tissue. We selected 7 datasets from the Oncomine database for comparison of gene expression, including Hou (normal = 65 and lung adenocarcinoma = 45), Landi (normal = 49 and lung adenocarcinoma = 58), Selamat (normal = 58 and lung adenocarcinoma = 58), Okayama (normal = 20 and lung adenocarcinoma = 226), Su (normal = 30 and lung adenocarcinoma = 27), Wei (normal = 25 and lung adenocarcinoma = 25), and Stearman (normal = 19 and lung adenocarcinoma = 20). The heatmap analysis indicated that the genetic expression patterns of 17 genes were similar among these datasets (Figure 2), which suggests that this molecular change is consistent in lung adenocarcinoma, and may represent a novel genetic signature in lung adenocarcinoma.

We classified 17 differentially expressed genes into 6 groups by biological and molecular functions, based on literature searches (Table 2; Supplementary Table 1), including (1) transcription regulation – TOX3 and SPDEF, (2) metabolism – PDK4, FABP4, and FMO2, (3) cell cycle regulation – CDKN2A, PHLD2A, SFN, and NDRG4, (4) cellular migration – AGR2, AQP5, and CLDN3 (5) inflammation – IL33, and (6) undefined group – ZDHHC9, BTLN9, GNG11, and CPED1 (also known as C7orF58). To further elucidate the role of these genetic expression changes in cancer progression, we performed survival curve analysis using the PrognoScan, Kaplan–Meier plotter, and SurvExpress databases.

Analysis of TOX3 and SPDEF in lung adenocarcinoma

The mRNA expression of TOX3 and SPDEF between lung adenocarcinoma and normal lung tissue derived from the Oncomine database is listed in Table 3. The analysis criteria were fold change > 2, p-value < 1E-04, and gene ranking in the top 10%. The results showed that both TOX3 and SPDEF are significantly upregulated in lung adenocarcinoma, compared to normal tissue. In addition, we selected a microarray with the accession number of GSE10072 from the Gene Expression Omnibus (GEO) database for gene expression analysis. This array contains 31 pairs of clinical lung adenocarcinomas and adjacent normal tissue. The results
showed that the mRNA expression of either TOX3 (Figure 3A) or SPDEF (Figure 3B) is upregulated in lung adenocarcinoma. To further investigate the role of TOX3 and SPDEF expression in cancer progression, we performed a survival curve analysis to evaluate the effects of gene expression in lung cancer patients with lung adenocarcinoma. The results indicated that the population with higher TOX3 expression has better survival rates (Figure 3C–3G), whereas the population with higher SPDEF expression has poorer survival outcome (Figure 3H, 3I). The prognostic values of TOX3 and SPDEF expression in lung adenocarcinoma were shown as forest plots (Figure 3J), which were derived from the PrognoScan database with a Cox p-value < 0.05, and the Kaplan–Meier plotter database with a log-rank p-value < 0.05.

Analysis of PDK4, FMO2, and FABP4 in lung adenocarcinoma

mRNA expression of PDK4, FMO2, and FABP4 between lung adenocarcinoma and normal lung tissue was analyzed by using the Oncomine database, and listed in Table 4. The expressions of PDK4, FMO2, and FABP4 are significantly downregulated in lung adenocarcinoma compared to normal lung tissue in several datasets, and this phenomenon was also observed in GSE10072 array (Figure 4A–4C). In cancer patients with lung

Figure 1: Identification of differentially expressed genes in lung adenocarcinoma compared to adjacent normal tissue using next-generation sequencing. Venn diagram analysis showed 9 upregulated genes (A) and 8 downregulated genes (B) in lung adenocarcinoma, compared to adjacent non-adenocarcinoma tissue from 3 pairs of clinical specimens. The criteria were fold change > 2 (tumor/normal) and fragments per kilobase million (FPKM) > 0.3. (C) The heatmap diagram showed the differentially expressed genes with z-score (log2) values by using color clustering on GENE-E web-tool. Green represents downregulation (minimum = –2.5), and red represents upregulation (maximum = 2.5).

www.impactjournals.com/oncotarget 104833 Oncotarget
adenocarcinoma, the survival curve analysis indicated that the population with higher expression of PDK4 (Figure 4D–4F), FMO2 (Figure 4G–4I), or FABP4 (Figure 4J) is correlated with better survival rates. The prognostic values of PDK4, FMO2, and FABP4 are shown in Figure 4K.

Analysis of CDKN2A, NDRG4, SFN, and PHLDA2 in lung adenocarcinoma

We observed that the expression of CDKN2A, PHLDA2, and SFN are upregulated and the expression of NDRG4 is downregulated in lung adenocarcinoma when compared to normal lung tissue, which was also confirmed by the Oncomine database, listed in Table 5. In the GSE10072 array, expression levels of CDKN2A, PHLDA2, and SFN are significantly upregulated in lung adenocarcinoma compared to normal lung tissue in several datasets (Figure 5A–5C), and NDRG4 is downregulated (Figure 5D). The survival curve analysis showed that lung adenocarcinoma patients with high levels of CDKN2A expression are associated with poorer survival rates (Figure 5E). The expression level of NDRG4, however, has no significant effects on survival outcomes for patients with lung adenocarcinoma (Figure 5F). Higher expressions of PHLDA2 (Figure 5G, 5H), or SFN (Figure 5I–5K) are also associated with poorer survival. The forest plots showed the prognostic values of CDKN2A, PHLDA2, SFN, and NDRG4 (Figure 5L).

Analysis of AGR2, CLDN3, and AQP5 in lung adenocarcinoma

Analysis of mRNA expression from the Oncomine database revealed that AGR2, AQP5, and CLDN3 are upregulated in lung adenocarcinoma, compared to normal lung tissue, and these results are listed in Table 6. In the GSE10072 array, we also observed that expression levels of AGR2 and CLDN3 are significantly upregulated in lung adenocarcinoma when compared to normal lung tissue (Figure 6A, 6B). However, the expression of AQP5 showed no significant change in the GSET10072 array (Figure 6C). The survival curve analysis showed that high expression of CLDN3 is correlated with poorer rates of survival in lung adenocarcinoma patients (Figure 6D). However, the population with a higher expression of AQP5 has better survival outcome (Figure 6E–6G). Higher expression of AGR2 is also associated with poorer survival rates (Figure 6H). The prognostic values of AGR2, CLDN3, and AQP5 were shown as forest plots (Figure 6I).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>FPKM (fragments per kilobase million)</th>
<th>T/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOX3</td>
<td>TOX High Mobility Group Box Family Member 3</td>
<td>1.65 11.99 0.70 11.57 1.08 31.11</td>
<td>Up</td>
</tr>
<tr>
<td>AGR2</td>
<td>Anterior gradient 2, Protein Disulphide Isomerase Family Member</td>
<td>340.96 1530.74 126.15 864.49 148.26 955.51</td>
<td>Up</td>
</tr>
<tr>
<td>SPDEF</td>
<td>SAM Pointed Domain Containing ETS Transcription Factor</td>
<td>5.11 37.10 0.48 3.91 0.88 5.90</td>
<td>Up</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Cyclin-Dependent Kinase Inhibitor 2A</td>
<td>7.04 298.07 10.05 129.69 4.94 80.09</td>
<td>Up</td>
</tr>
<tr>
<td>AQP5</td>
<td>Aquaporin 5</td>
<td>41.76 333.28 12.70 59.02 14.62 107.83</td>
<td>Up</td>
</tr>
<tr>
<td>CLDN3</td>
<td>Claudin 3</td>
<td>24.83 104.71 6.02 38.53 7.31 39.11</td>
<td>Up</td>
</tr>
<tr>
<td>SFN</td>
<td>Stratifin</td>
<td>27.16 75.72 7.48 59.63 9.59 65.88</td>
<td>Up</td>
</tr>
<tr>
<td>PHLDA2</td>
<td>Pleckstrin Homology-Like Domain, Family A, Member 2</td>
<td>9.07 96.14 2.60 19.93 2.36 18.58</td>
<td>Up</td>
</tr>
<tr>
<td>ZDHHC9</td>
<td>Zinc finger, DHHC-type containing 9</td>
<td>17.89 53.14 12.38 49.49 11.08 78.66</td>
<td>Up</td>
</tr>
<tr>
<td>PDK4</td>
<td>Pyruvate Dehydrogenase Kinase 4</td>
<td>30.91 10.82 72.24 3.18 47.76 7.63</td>
<td>Down</td>
</tr>
<tr>
<td>FMO2</td>
<td>Flavin Containing Monoxygenase 2</td>
<td>50.62 13.47 138.82 11.57 185.08 15.11</td>
<td>Down</td>
</tr>
<tr>
<td>NDRG4</td>
<td>NDRG Family Member 4</td>
<td>13.43 3.88 40.80 5.08 26.46 6.07</td>
<td>Down</td>
</tr>
<tr>
<td>CPED1</td>
<td>Cadherin-like and PC-esterase Domain Containing 1</td>
<td>16.73 5.70 19.42 5.68 23.24 5.39</td>
<td>Down</td>
</tr>
<tr>
<td>GNG11</td>
<td>G Protein Subunit Gamma 11</td>
<td>32.96 11.15 57.85 5.37 42.70 12.85</td>
<td>Down</td>
</tr>
<tr>
<td>IL33</td>
<td>Interleukin 33</td>
<td>117.77 33.34 133.98 16.49 153.31 15.49</td>
<td>Down</td>
</tr>
<tr>
<td>BTNL9</td>
<td>Butyrophilin-like 9</td>
<td>19.57 6.43 32.15 0.92 26.15 5.25</td>
<td>Down</td>
</tr>
<tr>
<td>FABP4</td>
<td>Fatty Acid Binding Protein 4</td>
<td>252.56 74.90 486.96 5.08 459.21 49.26</td>
<td>Down</td>
</tr>
</tbody>
</table>
### Table 2: Functional classification of differentially expressed genes

<table>
<thead>
<tr>
<th>Functions</th>
<th>Genes</th>
<th>Fold change (Cancer/Normal)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcriptional regulation</td>
<td>TOX3</td>
<td>UP</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>SPDEF</td>
<td>UP</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>PDK4</td>
<td>DOWN</td>
<td>3</td>
</tr>
<tr>
<td>Metabolism</td>
<td>FABP4</td>
<td>DOWN</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>FMO2</td>
<td>DOWN</td>
<td>5, 6</td>
</tr>
<tr>
<td></td>
<td>CDKN2A</td>
<td>UP</td>
<td>7</td>
</tr>
<tr>
<td>Cell cycle regulation</td>
<td>PHLDA2</td>
<td>UP</td>
<td>8, 9</td>
</tr>
<tr>
<td></td>
<td>NDRG4</td>
<td>DOWN</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>SFN</td>
<td>UP</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>AGR2</td>
<td>UP</td>
<td>12</td>
</tr>
<tr>
<td>Cellular migration</td>
<td>AQP5</td>
<td>UP</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>CLDN3</td>
<td>UP</td>
<td>14</td>
</tr>
<tr>
<td>Inflammation</td>
<td>IL33</td>
<td>DOWN</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>ZDHHC9</td>
<td>UP</td>
<td>16</td>
</tr>
<tr>
<td>Others</td>
<td>BTNL9</td>
<td>DOWN</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>GNG11</td>
<td>DOWN</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>CPED1</td>
<td>DOWN</td>
<td>19</td>
</tr>
</tbody>
</table>

* References were listed in Supplementary Table 1.

### Table 3: Analysis of TOX3 and SPDEF mRNA expression in lung adenocarcinoma compared to normal tissue from Oncomine database

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change (Cancer/Normal)</th>
<th>P-value</th>
<th>Gene Ranking (Top%)</th>
<th>Samples (Normal : Tumor)</th>
<th>Dataset</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOX3</td>
<td>8.685</td>
<td>1.18E-22</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>215108_x_at</td>
</tr>
<tr>
<td></td>
<td>19.973</td>
<td>1.86E-24</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>214774_x_at</td>
</tr>
<tr>
<td></td>
<td>4.968</td>
<td>2.80E-18</td>
<td>2</td>
<td>49 : 58</td>
<td>Landi</td>
<td>216623_x_at</td>
</tr>
<tr>
<td></td>
<td>8.497</td>
<td>1.79E-9</td>
<td>1</td>
<td>30 : 27</td>
<td>Su</td>
<td>216623_x_at</td>
</tr>
<tr>
<td></td>
<td>3.15</td>
<td>3.09E-4</td>
<td>2</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>37426_at</td>
</tr>
<tr>
<td></td>
<td>12.617</td>
<td>2.17E-7</td>
<td>3</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>37426_at</td>
</tr>
<tr>
<td></td>
<td>3.764</td>
<td>4.72E-7</td>
<td>10</td>
<td>65 : 45</td>
<td>Hou</td>
<td>216623_x_at</td>
</tr>
<tr>
<td></td>
<td>3.719</td>
<td>1.54E-11</td>
<td>5</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>213441_x_at</td>
</tr>
<tr>
<td></td>
<td>3.451</td>
<td>2.64E-12</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>220192_x_at</td>
</tr>
<tr>
<td></td>
<td>5.44</td>
<td>2.86E-9</td>
<td>1</td>
<td>30 : 27</td>
<td>Su</td>
<td>220192_x_at</td>
</tr>
<tr>
<td></td>
<td>2.12</td>
<td>1.94E-7</td>
<td>9</td>
<td>65 : 45</td>
<td>Hou</td>
<td>220192_x_at</td>
</tr>
<tr>
<td></td>
<td>3.844</td>
<td>9.6E-17</td>
<td>2</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_2161330</td>
</tr>
<tr>
<td>SPDEF</td>
<td>6.091</td>
<td>1.03E-13</td>
<td>1</td>
<td>34 : 18</td>
<td>Okayama</td>
<td>213008_x_at</td>
</tr>
<tr>
<td></td>
<td>11.722</td>
<td>3.80E-13</td>
<td>1</td>
<td>33 : 24</td>
<td>Wei</td>
<td>220192_x_at</td>
</tr>
<tr>
<td></td>
<td>7.775</td>
<td>1.13E-16</td>
<td>1</td>
<td>54 : 12</td>
<td>Selamat</td>
<td>ILMN_2161330</td>
</tr>
<tr>
<td></td>
<td>3.16</td>
<td>1.32E-4</td>
<td>1</td>
<td>50 : 30</td>
<td>Sou</td>
<td>213441_x_at</td>
</tr>
</tbody>
</table>

www.impactjournals.com/oncotarget 104835 Oncotarget
Analysis of IL33 in lung adenocarcinoma

The mRNA expression of IL33 is significantly downregulated in lung adenocarcinoma compared to normal tissue (Table 7). We also found that IL33 expression is decreased in lung adenocarcinoma identified from the GSE10072 array (Figure 7A). The survival curve analysis performed using the SurvExpress database showed that the high risk population with lower expression of IL33 has poorer survival outcomes for lung adenocarcinoma patients (Figure 7B, 7C). This phenomenon was also observed in the PrognoScan (Figure 7D–7F) and Kaplan–Meier plotter databases (Figure 7G). The prognostic values of IL33 in lung adenocarcinoma was illustrated as a forest plot (Figure 7H).

Analysis of ZDHHC9, BTNL9, GNG11, and CPED1 in lung adenocarcinoma

The mRNA expression of BTNL9, GNG11, or CPED1 is downregulated in lung adenocarcinoma when compared to normal tissue, and ZDHHC9 is upregulated (Table 8). In the GSE10072 array, GNG11 (Figure 8A)

Figure 2: Comparison of differentially expressed genes in clinical lung adenocarcinoma and normal lung tissue by Oncomine database analysis. Seven microarray datasets from the Oncomine database were used to analyze gene expression patterns (lung adenocarcinoma vs. normal), including (A) Hou, (B) Landi, (C) Selamat, (D) Okayama, (E) Su, (F) Wei, and (G) Stearman. Seventeen differentially expressed genes (9 up and 8 down) identified from 3 pairs of clinical lung adenocarcinoma were selected. Raw data were extracted and re-plotted by GENE-E web-tool, and the relative color scheme used for clustering analysis. Yellow represents high expression (maximum = 1) and blue represents low expression (minimum = 0). The gene symbols and corresponding specific probes are displayed on the right side of each diagram.
Figure 3: Analysis of TOX3 and SPDEF in clinical lung adenocarcinoma patients using bioinformatics databases.

The gene expression of TOX3 (A) and SPDEF (B), comparing 31 pairs of clinical lung adenocarcinoma (red) and adjacent normal tissue (blue), was performed on GSE10072 microarray from the GEO database. (3 probes for TOX3 and SPDEF respectively in GSE10072) The p-value of gene expression was calculated by t-test with Wilcoxon matched-pairs signed rank test. *** represents p < 0.001, ** represents p < 0.01. The survival curves comparing 2 populations with high (red) and low (black) gene expression in lung adenocarcinoma patients were performed on the PrognoScan database - TOX3 (C–F) and SPDEF (H, I), and the Kaplan–Meier plotter database - TOX3 (G). The analysis criteria of the PrognoScan and Kaplan–Meier plotter databases were Cox p-value < 0.05 and log-rank p-value < 0.05 respectively.

The raw data of GEO and PrognoScan databases were extracted and re-plotted by GraphPad Prism 5 software. (J) The forest plots showed hazard ratios (95% CI, confidence interval) identified from PrognoScan and Kaplan–Meier plotter databases.
and CPED1 (Figure 8B) expression is downregulated in lung adenocarcinoma. There is no specific probe for ZDHHC9 and BTNL9 in the GSE10072 array. Survival curve analysis revealed that lung adenocarcinoma patients with high expression of ZDHHC9 (Figure 8C), BTNL9 (Figure 8D–8F), GNG11 (Figure 8G, 8H) or CPED1 (Figure 8I–8K) are correlated with better survival outcomes. The forest plot showed the prognostic values of ZDHHC9, BTNL9, GNG11, and CPED1 (Figure 8L).

**DISCUSSION**

Lung cancer, one of the leading causes of cancer-related death worldwide [29], still has much that requires

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change (Cancer/Normal)</th>
<th>P-value</th>
<th>Gene Ranking (Top%)</th>
<th>Samples (Normal : Tumor)</th>
<th>Dataset</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDK4</td>
<td>-5.162</td>
<td>1.27E-24</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>225207_at</td>
</tr>
<tr>
<td></td>
<td>-9.834</td>
<td>2.27E-13</td>
<td>2</td>
<td>25 : 25</td>
<td>Wei</td>
<td>225207_at</td>
</tr>
<tr>
<td></td>
<td>-2.415</td>
<td>1.92E-16</td>
<td>4</td>
<td>49 : 58</td>
<td>Landi</td>
<td>205960_at</td>
</tr>
<tr>
<td></td>
<td>-4.074</td>
<td>2.98E-7</td>
<td>5</td>
<td>30 : 27</td>
<td>Su</td>
<td>205960_at</td>
</tr>
<tr>
<td></td>
<td>-4.33</td>
<td>1.56E-4</td>
<td>8</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>36739_at</td>
</tr>
<tr>
<td></td>
<td>-2.618</td>
<td>1.77E-4</td>
<td>9</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>36739_at</td>
</tr>
<tr>
<td></td>
<td>-4.593</td>
<td>6.29E-27</td>
<td>2</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1684982</td>
</tr>
<tr>
<td></td>
<td>-3.095</td>
<td>9.03E-7</td>
<td>7</td>
<td>10 : 86</td>
<td>Beer</td>
<td>ILMN_1684982</td>
</tr>
<tr>
<td></td>
<td>-10.368</td>
<td>2.77E-6</td>
<td>2</td>
<td>5 : 40</td>
<td>Garber</td>
<td>IMAGE:78946</td>
</tr>
<tr>
<td></td>
<td>-3.77</td>
<td>4.98E-18</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>211726_s_at</td>
</tr>
<tr>
<td></td>
<td>-7.811</td>
<td>4.49E-15</td>
<td>2</td>
<td>25 : 25</td>
<td>Wei</td>
<td>228268_at</td>
</tr>
<tr>
<td></td>
<td>-5.528</td>
<td>6.60E-28</td>
<td>1</td>
<td>49 : 58</td>
<td>Landi</td>
<td>211726_s_at</td>
</tr>
<tr>
<td></td>
<td>-4.329</td>
<td>2.47E-10</td>
<td>2</td>
<td>30 : 27</td>
<td>Su</td>
<td>211726_s_at</td>
</tr>
<tr>
<td></td>
<td>-8.705</td>
<td>1.62E-53</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1732158</td>
</tr>
<tr>
<td></td>
<td>-9.102</td>
<td>1.02E-24</td>
<td>1</td>
<td>65 : 65</td>
<td>Hou</td>
<td>228268_at</td>
</tr>
<tr>
<td></td>
<td>-7.56</td>
<td>7.38E-6</td>
<td>2</td>
<td>5 : 39</td>
<td>Garber</td>
<td>IMAGE:8057</td>
</tr>
<tr>
<td></td>
<td>-11.062</td>
<td>9.47E-19</td>
<td>1</td>
<td>10 : 86</td>
<td>Beer</td>
<td>Y09267_at</td>
</tr>
<tr>
<td></td>
<td>-14.421</td>
<td>9.74E-28</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>203980_at</td>
</tr>
<tr>
<td></td>
<td>-20.112</td>
<td>1.42E-17</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>203980_at</td>
</tr>
<tr>
<td></td>
<td>-19.625</td>
<td>1.89E-17</td>
<td>1</td>
<td>30 : 27</td>
<td>Su</td>
<td>203980_at</td>
</tr>
<tr>
<td></td>
<td>-14.293</td>
<td>1.90E-37</td>
<td>1</td>
<td>49 : 58</td>
<td>Landi</td>
<td>203980_at</td>
</tr>
<tr>
<td>FMO2</td>
<td>-68.043</td>
<td>3.67E-13</td>
<td>1</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>38430_at</td>
</tr>
<tr>
<td></td>
<td>-13.918</td>
<td>2.07E-7</td>
<td>4</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>38430_at</td>
</tr>
<tr>
<td></td>
<td>-26.532</td>
<td>1.26E-12</td>
<td>2</td>
<td>10 : 86</td>
<td>Beer</td>
<td>Image:8057</td>
</tr>
<tr>
<td></td>
<td>-9.214</td>
<td>1.34E-5</td>
<td>3</td>
<td>5 : 40</td>
<td>Garber</td>
<td>IMAGE:2308848</td>
</tr>
<tr>
<td></td>
<td>-12.842</td>
<td>5.80E-44</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>IMAGE:2308848</td>
</tr>
</tbody>
</table>

and CPED1 (Figure 8B) expression is downregulated in lung adenocarcinoma. There is no specific probe for ZDHHC9 and BTNL9 in the GSE10072 array. Survival curve analysis revealed that lung adenocarcinoma patients with high expression of ZDHHC9 (Figure 8C), BTNL9 (Figure 8D–8F), GNG11 (Figure 8G, 8H) or CPED1 (Figure 8I–8K) are correlated with better survival outcomes. The forest plot showed the prognostic values of ZDHHC9, BTNL9, GNG11, and CPED1 (Figure 8L).

**Identification of genetic regulation in lung adenocarcinoma using next-generation sequencing**

We simultaneously performed small RNA-seq in these 3 pairs of specimens using next-generation sequencing (Figure 9). We focused on microRNAs and found 22 upregulated microRNAs in lung adenocarcinoma using Venn diagram analysis (Figure 9A), which is listed in Table 9. The analysis criteria were fold change > 2 and reads per million (RPM) > 1. No microRNA with downregulated changes were found in our analysis (Figure 9B). Heatmap color clustering showed the expression patterns of each upregulated microRNA from these 3 pairs of specimens (Figure 9C). To further elucidate the genetic interactions in lung adenocarcinoma, we performed miRmap database for target prediction. Among 22 upregulated microRNAs, there were 13 putative targets, shown in the “Targets” Venn diagram (Figure 9D). The prediction threshold was miRmap score > 90.0. The Venn diagram analysis between 13 targets of microRNAs and 8 downregulated genes, shown in Figure 1B, indicates that there were 10 genetic interactions of microRNA-mRNA in lung adenocarcinoma (Figure 9D), which is listed in Table 10. Only 6 genes were involved in these 10 genetic interactions, due to the 3 genes have been attributed to different microRNAs.

**DISCUSSION**

Lung cancer, one of the leading causes of cancer-related death worldwide [29], still has much that requires
Figure 4: Analysis of PDK4, FMO2 and FABP4 in clinical lung adenocarcinoma patients using bioinformatics databases. The gene expression of PDK4 (A), FMO2 (B), and FABP4 (C), comparing 31 pairs of clinical lung adenocarcinoma (red) and adjacent normal tissue (blue), was performed on GSE10072 microarray from the GEO database. The p-value of gene expression was calculated by t-test with Wilcoxon matched-pairs signed rank test. *** represents p < 0.001. The survival curves comparing 2 populations with high (red) and low (black) gene expression in lung adenocarcinoma patients were performed on the PrognoScan database - PDK4 (D, E), FMO2 (G), and FABP4 (J), and the Kaplan–Meier plotter database – PDK4 (F) and FMO2 (H, I). The analysis criteria of the PrognoScan and Kaplan–Meier plotter databases were Cox p-value < 0.05 and log-rank p-value < 0.05 respectively. Raw data of the GEO and the PrognoScan databases were extracted and re-plotted by GraphPad Prism 5 software. (K) The forest plots showed hazard ratios (95% CI, confidence interval) identified from the PrognoScan and Kaplan–Meier plotter databases.
In our project, we hoped to identify novel gene expression signature or genetic interactions of gene-microRNA in lung adenocarcinoma by using next-generation sequencing combined with systematic bioinformatics analysis.

We found 17 differentially expressed genes in lung adenocarcinoma compared to its adjacent normal lung tissue, which were classified into 6 functional groups based on a search of the literature. These results indicated that tumor progression is involved in alterations of various biological functions. We then summarized the potential oncogenic and tumor suppressor roles of these genes in lung adenocarcinoma (Supplementary Table 2).

Further study. In our project, we hoped to identify novel gene expression signature or genetic interactions of gene-microRNA in lung adenocarcinoma by using next-generation sequencing combined with systematic bioinformatics analysis.

We found 17 differentially expressed genes in lung adenocarcinoma compared to its adjacent normal lung tissue, which were classified into 6 functional groups based on a search of the literature. These results indicated that tumor progression is involved in alterations of various biological functions. We then summarized the potential oncogenic and tumor suppressor roles of these genes in lung adenocarcinoma (Supplementary Table 2).

TOX3 contains an HMG-box (high mobility group box) domain. The function of TOX3 remains unclear, but it may be involved in various DNA-dependent processes [30–32]. TOX3 polymorphisms and epigenetic regulation have been demonstrated in breast cancer [33] and lung cancer [34] respectively. In our study, TOX3 was significantly upregulated in lung adenocarcinoma, and higher expression of TOX3 is correlated with better survival outcome. We speculated that as more factors may be involved in TOX3-related mechanisms of tumor progression, more studies are needed to clarify the relationship between TOX3 expression and tumor progression.

SPDEF containing ETS domain has been reported to be overexpressed in many cancers [35–37]. Our study suggests that SPDEF may play an oncogenic role in lung adenocarcinoma, although some reports have shown that SPDEF can suppress cancer metastasis [38]. However, contrary effects of SPDEF on tumorigenesis require further research.

PDK4 is a mitochondrial protein that can regulate glucose metabolism through inhibition of pyruvate dehydrogenase complex. An aberrant metabolism is one of the characteristics of cancer cells. In liver cancer, PDK4 has been identified as a potential tumor suppressor [39].

### Table 5: Analysis of CDKN2A, PHLDA2, SFN, and NDRG4 mRNA expression in lung adenocarcinoma compared to normal tissue from Oncomine database

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change (Cancer/Normal)</th>
<th>P-value</th>
<th>Gene Ranking (Top%)</th>
<th>Samples (Normal : Tumor)</th>
<th>Dataset</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN2A</td>
<td>2.793</td>
<td>1.08E-8</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>225207_at</td>
</tr>
<tr>
<td></td>
<td>3.203</td>
<td>2.43E-6</td>
<td>2</td>
<td>25 : 25</td>
<td>Wei</td>
<td>225207_at</td>
</tr>
<tr>
<td></td>
<td>2.030</td>
<td>7.82E-11</td>
<td>4</td>
<td>49 : 58</td>
<td>Landi</td>
<td>205960_at</td>
</tr>
<tr>
<td></td>
<td>3.161</td>
<td>4.32E-9</td>
<td>5</td>
<td>65 : 45</td>
<td>Hou</td>
<td>205960_at</td>
</tr>
<tr>
<td></td>
<td>2.206</td>
<td>0.01</td>
<td>8</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>36739_at</td>
</tr>
<tr>
<td></td>
<td>3.506</td>
<td>7.68E-4</td>
<td>9</td>
<td>10 : 86</td>
<td>Beer</td>
<td>36739_at</td>
</tr>
<tr>
<td></td>
<td>4.207</td>
<td>4.25E-9</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>211726_s_at</td>
</tr>
<tr>
<td></td>
<td>5.73</td>
<td>1.80E-4</td>
<td>2</td>
<td>17 : 132</td>
<td>Bhattacharyee</td>
<td>228268_at</td>
</tr>
<tr>
<td></td>
<td>4.027</td>
<td>1.99E-19</td>
<td>1</td>
<td>49 : 58</td>
<td>Landi</td>
<td>211726_s_at</td>
</tr>
<tr>
<td></td>
<td>3.634</td>
<td>1.09E-7</td>
<td>2</td>
<td>30 : 27</td>
<td>Su</td>
<td>211726_s_at</td>
</tr>
<tr>
<td></td>
<td>2.387</td>
<td>1.42E-17</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1732158</td>
</tr>
<tr>
<td></td>
<td>2.212</td>
<td>1.35E-5</td>
<td>1</td>
<td>65 : 45</td>
<td>Hou</td>
<td>228268_at</td>
</tr>
<tr>
<td></td>
<td>4.015</td>
<td>9.11E-8</td>
<td>2</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>IMAGE:80507</td>
</tr>
<tr>
<td></td>
<td>2.799</td>
<td>2.05E-11</td>
<td>1</td>
<td>10 : 86</td>
<td>Beer</td>
<td>Y09267_at</td>
</tr>
<tr>
<td></td>
<td>4.721</td>
<td>1.98E-11</td>
<td>2</td>
<td>25 : 25</td>
<td>Wei</td>
<td>33323_r_at</td>
</tr>
<tr>
<td></td>
<td>2.049</td>
<td>1.60E-12</td>
<td>5</td>
<td>49 : 58</td>
<td>Landi</td>
<td>33322_i_at</td>
</tr>
<tr>
<td></td>
<td>6.459</td>
<td>3.04E-8</td>
<td>2</td>
<td>30 : 27</td>
<td>Su</td>
<td>209260_at</td>
</tr>
<tr>
<td>SFN</td>
<td>2.410</td>
<td>4.06E-6</td>
<td>5</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>33322_i_at</td>
</tr>
<tr>
<td></td>
<td>4.487</td>
<td>1.67E-24</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1806607</td>
</tr>
<tr>
<td></td>
<td>2.302</td>
<td>2.74E-5</td>
<td>6</td>
<td>10 : 86</td>
<td>Beer</td>
<td>X57348_s_at</td>
</tr>
<tr>
<td></td>
<td>-3.519</td>
<td>1.78E-12</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>203980_at</td>
</tr>
<tr>
<td></td>
<td>-5.009</td>
<td>1.03E-13</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>203980_at</td>
</tr>
<tr>
<td></td>
<td>-3.923</td>
<td>5.21E-9</td>
<td>1</td>
<td>30 : 27</td>
<td>Su</td>
<td>203980_at</td>
</tr>
<tr>
<td>NDRG4</td>
<td>-4.402</td>
<td>2.32E-7</td>
<td>4</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>38430_at</td>
</tr>
<tr>
<td></td>
<td>-2.448</td>
<td>1.60E-14</td>
<td>2</td>
<td>65 : 45</td>
<td>Hou</td>
<td>J02874_at</td>
</tr>
<tr>
<td></td>
<td>-2.152</td>
<td>4.63E-4</td>
<td>3</td>
<td>5 : 40</td>
<td>Garber</td>
<td>IMAGE:2308848</td>
</tr>
<tr>
<td></td>
<td>-3.007</td>
<td>1.04E-18</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1773006</td>
</tr>
</tbody>
</table>

Further study. In our project, we hoped to identify novel gene expression signature or genetic interactions of gene-microRNA in lung adenocarcinoma by using next-generation sequencing combined with systematic bioinformatics analysis.

We found 17 differentially expressed genes in lung adenocarcinoma compared to its adjacent normal lung tissue, which were classified into 6 functional groups based on a search of the literature. These results indicated that tumor progression is involved in alterations of various biological functions. We then summarized the potential oncogenic and tumor suppressor roles of these genes in lung adenocarcinoma (Supplementary Table 2).

TOX3 contains an HMG-box (high mobility group box) domain. The function of TOX3 remains unclear, but it may be involved in various DNA-dependent processes [30–32]. TOX3 polymorphisms and epigenetic regulation have been demonstrated in breast cancer [33] and lung cancer [34] respectively. In our study, TOX3 was significantly upregulated in lung adenocarcinoma, and higher expression of TOX3 is correlated with better survival outcome. We speculated that as more factors may be involved in TOX3-related mechanisms of tumor progression, more studies are needed to clarify the relationship between TOX3 expression and tumor progression.

SPDEF containing ETS domain has been reported to be overexpressed in many cancers [35–37]. Our study suggests that SPDEF may play an oncogenic role in lung adenocarcinoma, although some reports have shown that SPDEF can suppress cancer metastasis [38]. However, contrary effects of SPDEF on tumorigenesis require further research.

PDK4 is a mitochondrial protein that can regulate glucose metabolism through inhibition of pyruvate dehydrogenase complex. An aberrant metabolism is one of the characteristics of cancer cells. In liver cancer, PDK4 has been identified as a potential tumor suppressor [39].
Figure 5: Analysis of CDKN2A, PHLDA2, SFN, and NDRG4 in clinical lung adenocarcinoma patients using bioinformatics databases. The gene expression of CDKN2A (A), PHLDA2 (B), SFN (C), and NDRG4 (D), comparing 31 pairs of clinical lung adenocarcinoma (red) and adjacent normal tissue (blue), was performed on a GSE10072 microarray from the GEO database. The p-value of gene expression was calculated by t-test with Wilcoxon matched-pairs signed rank test. *** represents p < 0.001, ** represents p < 0.01. The survival curves comparing 2 populations with high (red) and low (black) gene expression in lung adenocarcinoma patients were performed on the Kaplan–Meier plotter database – CDKN2A (E) and NDRG4 (F), and PrognoScan database – PHLDA2 (G, H) and SFN (I–K). The analysis criteria of PrognoScan and Kaplan–Meier plotter databases were Cox p-value < 0.05 and log-rank p-value < 0.05 respectively. Raw data of the GEO and PrognoScan databases were extracted and re-plotted by GraphPad Prism 5 software. (L) The forest plots showed hazard ratios (95% CI, confidence interval) identified from the PrognoScan and Kaplan–Meier plotter databases.
Ironically, PDK4 exerts oncogenic effects in colon cancer [40]. In our study, PDK4 played a potential tumor suppressor role in lung adenocarcinoma.

FMO2 is an NADPH-dependent enzyme that catalyzes the oxygenation of substrates [41], but the effect of FMO2 on tumorigenesis is unclear. Although genetic polymorphisms of FMO genes may influence drug metabolism [42], we found that FMO2 might have tumor suppressor effects in lung adenocarcinoma.

FABP4 is involved in fatty acids trafficking and metabolism. Fatty acids serve as both an energy source and signaling molecules that can regulate various cellular functions [43]. The dysfunction of FABP proteins has been found to be associated with some metabolic diseases [44], and elevated FABP4 has been observed in many types of cancer [45–47]. Our data showed that FABP4 may have tumor suppressor effects in lung adenocarcinoma.

CDKN2A encodes two spliced transcripts, p16INK4a and p14ARF, which regulate cell cycle progression through inhibition of CDK4 kinase and p53 respectively [48]. CDKN2A has been shown as a tumor suppressor in cancer progression [49], and its alterations, including epigenetic modifications, deletion, and mutations, frequently occur in cancers [50].

### Table 6: Analysis of AGR2, AQP5, and CLDN3 mRNA expression in lung adenocarcinoma compared to normal tissue from Oncomine database

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change (Cancer/Normal)</th>
<th>P-value</th>
<th>Gene Ranking (Top%)</th>
<th>Samples (Normal : Tumor)</th>
<th>Dataset</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGR2</td>
<td>2.965</td>
<td>9.38E-11</td>
<td>6</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>228969_at</td>
</tr>
<tr>
<td></td>
<td>5.586</td>
<td>1.83E-15</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>209173_at</td>
</tr>
<tr>
<td></td>
<td>2.779</td>
<td>9.67E-11</td>
<td>7</td>
<td>49 : 58</td>
<td>Landi</td>
<td>209173_at</td>
</tr>
<tr>
<td></td>
<td>3.434</td>
<td>1.49E-8</td>
<td>1</td>
<td>30 : 27</td>
<td>Su</td>
<td>209173_at</td>
</tr>
<tr>
<td></td>
<td>2.902</td>
<td>0.003</td>
<td>3</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>38827_at</td>
</tr>
<tr>
<td></td>
<td>2.677</td>
<td>1.80E-5</td>
<td>6</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>38827_at</td>
</tr>
<tr>
<td></td>
<td>2.393</td>
<td>5.73E-11</td>
<td>7</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1814151</td>
</tr>
<tr>
<td>AQP5</td>
<td>2.941</td>
<td>2.80E-6</td>
<td>10</td>
<td>25 : 25</td>
<td>Wei</td>
<td>213611_at</td>
</tr>
<tr>
<td></td>
<td>3.287</td>
<td>5.33E-12</td>
<td>4</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>209354_x_at</td>
</tr>
<tr>
<td></td>
<td>3.594</td>
<td>1.50E-11</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>209354_x_at</td>
</tr>
<tr>
<td></td>
<td>3.282</td>
<td>1.71E-7</td>
<td>2</td>
<td>30 : 27</td>
<td>Su</td>
<td>209354_x_at</td>
</tr>
<tr>
<td>CLDN3</td>
<td>2.193</td>
<td>4.34E-16</td>
<td>2</td>
<td>49 : 58</td>
<td>Landi</td>
<td>209354_x_at</td>
</tr>
<tr>
<td></td>
<td>3.707</td>
<td>1.61E-8</td>
<td>1</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>33904_at</td>
</tr>
<tr>
<td></td>
<td>5.152</td>
<td>3.60E-5</td>
<td>1</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>33904_at</td>
</tr>
<tr>
<td></td>
<td>3.154</td>
<td>3.74E-14</td>
<td>4</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1723042</td>
</tr>
</tbody>
</table>

### Table 7: Analysis of IL33 mRNA expression in lung adenocarcinoma compared to normal tissue from Oncomine database

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change (Cancer/Normal)</th>
<th>P-value</th>
<th>Gene Ranking (Top%)</th>
<th>Samples (Normal : Tumor)</th>
<th>Dataset</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL33</td>
<td>–3.809</td>
<td>6.11E-20</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>209821_at</td>
</tr>
<tr>
<td></td>
<td>–4.325</td>
<td>1.04E-9</td>
<td>5</td>
<td>25 : 25</td>
<td>Wei</td>
<td>209821_at</td>
</tr>
<tr>
<td></td>
<td>–3.276</td>
<td>1.23E-21</td>
<td>2</td>
<td>49 : 58</td>
<td>Landi</td>
<td>209821_at</td>
</tr>
<tr>
<td></td>
<td>–7.088</td>
<td>1.82E-9</td>
<td>3</td>
<td>30 : 27</td>
<td>Su</td>
<td>209821_at</td>
</tr>
<tr>
<td></td>
<td>–3.582</td>
<td>2.02E-11</td>
<td>7</td>
<td>65 : 45</td>
<td>Hou</td>
<td>209821_at</td>
</tr>
<tr>
<td></td>
<td>–2.163</td>
<td>8.36E-5</td>
<td>7</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>35333_r_at</td>
</tr>
<tr>
<td></td>
<td>–2.351</td>
<td>5.08E-5</td>
<td>8</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>35333_r_at</td>
</tr>
<tr>
<td></td>
<td>–4.258</td>
<td>5.82E-30</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1809099</td>
</tr>
</tbody>
</table>
Figure 6: Analysis of AGR2, CLDN3, and AQP5 in clinical lung adenocarcinoma patients using bioinformatics databases. The gene expression of AGR2 (A), CLDN3 (B), and AQP5 (C), comparing 31 pairs of clinical lung adenocarcinoma (red) and adjacent normal tissue (blue), was performed on a GSE10072 microarray from the GEO database. The p-value of gene expression was calculated by t-test with Wilcoxon matched-pairs signed rank test. *** represents p < 0.001 and n.s. represents no significance. Survival curves comparing 2 populations with high (red) and low (black) gene expression in lung adenocarcinoma patients were performed on the Kaplan–Meier plotter – CLDN3 (D) and AQP5 (E), and PrognoScan databases – AQP5 (F, G) and AGR2 (H). The analysis criteria of the PrognoScan and Kaplan–Meier plotter databases were Cox p-value < 0.05 and log-rank p-value < 0.05 respectively. The raw data of the GEO and PrognoScan databases were extracted and re-plotted by GraphPad Prism 5 software. (I) The forest plots showed hazard ratios (95% CI, confidence interval) identified from the PrognoScan and Kaplan–Meier plotter databases.
CDKN2A may have potential oncogenic effects in lung adenocarcinoma.

PHLDA2, located in an imprinted region on chromosome 11p15.5, has primarily been studied for its regulation of placental growth [51]. Although the role of PHLDA2 in cancer is unclear, our data showed that PHLDA2 may potentially exert oncogenic effects in lung adenocarcinoma.

SFN is involved in protein synthesis and epithelial cell growth. Numerous reports have demonstrated the molecular functions of SFN in keratinocytes and fibroblasts [52]. Furthermore, elevated expression of SFN

**Figure 7:** Analysis of IL33 in clinical lung adenocarcinoma patients using bioinformatics databases. The gene expression of IL33 (A), comparing 31 pairs of clinical lung adenocarcinoma (red) and adjacent normal tissue (blue), was performed on a GSE10072 microarray from the GEO database. The *p*-value of gene expression was calculated by *t*-test with Wilcoxon matched-pairs signed rank test. *** represents *p* < 0.001. (B) The survival curve was captured from the SurvExpress database, which divided lung adenocarcinoma patients into 2 populations of high (red) and low (green) risk, and (C) the box plots showed that the high risk (red) population has lower levels of IL33 expression, while the low risk (green) population has higher levels of IL33 expression. (D–F) The survival curves comparing 2 populations with high (red) and low (black) gene expression in lung adenocarcinoma patients were performed on the PrognoScan and (G) Kaplan–Meier plotter databases. The analysis criteria of the PrognoScan and Kaplan–Meier plotter databases were Cox *p*-value < 0.05 and log-rank *p*-value < 0.05 respectively. Raw data of the GEO and the PrognoScan databases were extracted and re-plotted by GraphPad Prism 5 software. (H) The forest plots showed hazard ratios (95% CI, confidence interval) identified from the PrognoScan and Kaplan–Meier plotter databases.
has been reported in lung adenocarcinoma [53], and our study also found that SFN may exert oncogenic effects in lung adenocarcinoma.

NDRG4 is involved in the regulation of cell cycle progression [54] and has been identified as a novel tumor suppressor in colon cancer [55]. We found that NDRG4 levels were significantly decreased in lung adenocarcinoma, but with regard to survival analysis, the expression of NDRG4 has shown no significant influence on survival rates of lung cancer patients.

AGR2 is an endoplasmic reticulum (ER) protein which can catalyze protein folding. Its oncogenic role and increased expression have been reported in different types of cancer [56–58]. In our study, we found that AGR2 may serve as a potential prognostic biomarker of lung adenocarcinoma.

AQP5, aquaporin 5, is a water channel protein involved in pulmonary secretions, and elevated expression of AQP5 is associated with poor survival outcome in many types of cancer [59–61]. In our study, we found that AQP5 was upregulated in lung adenocarcinoma, and its high expression correlated with better survival outcome.

CLDN3 regulates tight junctions of cell-cell adhesion in epithelial or endothelial cells and is overexpressed in ovarian [62] and colon cancer [63]. Loss of claudin 3 expression increases the metastatic ability of esophageal cancer [64], whereas claudin 3 is upregulated in lung adenocarcinoma [65]. Our study showed that claudin 3 may have oncogenic effects in lung adenocarcinoma.

IL33 is a cytokine involved in a spectrum of biological processes, and the chronic inflammatory signaling activation is known to be involved in cancer progression. The expression of IL33 in tumor tissues is depressed, but tumor stroma and serum have increased levels of IL33, suggesting the distinct functions of IL33 in cancer cells from the microenvironment [66]. IL-33 is shown to promote tumorigenesis and induce stemness in breast cancer [67]. In Apc

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change (Cancer/Normal)</th>
<th>P-value</th>
<th>Gene Ranking (Top%)</th>
<th>Samples (Normal : Tumor)</th>
<th>Dataset</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZDHHC9</td>
<td>2.295</td>
<td>8.84E-12</td>
<td>5</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>222451_s_at</td>
</tr>
<tr>
<td></td>
<td>2.621</td>
<td>9.16E-10</td>
<td>3</td>
<td>25 : 25</td>
<td>Wei</td>
<td>222451_s_at</td>
</tr>
<tr>
<td></td>
<td>2.377</td>
<td>3.27E-19</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1803824</td>
</tr>
<tr>
<td></td>
<td>–12.000</td>
<td>3.77E-14</td>
<td>3</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>228434_at</td>
</tr>
<tr>
<td>BTNL9</td>
<td>–10.916</td>
<td>6.99E-26</td>
<td>1</td>
<td>25 : 25</td>
<td>Wei</td>
<td>228434_at</td>
</tr>
<tr>
<td></td>
<td>–9.102</td>
<td>1.02E-24</td>
<td>1</td>
<td>65 : 45</td>
<td>Hou</td>
<td>228434_at</td>
</tr>
<tr>
<td></td>
<td>–3.630</td>
<td>9.97E-21</td>
<td>1</td>
<td>20 : 226</td>
<td>Okayama</td>
<td>204115_at</td>
</tr>
<tr>
<td></td>
<td>–4.337</td>
<td>6.55E-11</td>
<td>4</td>
<td>25 : 25</td>
<td>Wei</td>
<td>204115_at</td>
</tr>
<tr>
<td></td>
<td>–3.463</td>
<td>1.49E-12</td>
<td>1</td>
<td>30 : 27</td>
<td>Su</td>
<td>204115_at</td>
</tr>
<tr>
<td></td>
<td>–3.277</td>
<td>7.32E-23</td>
<td>2</td>
<td>49 : 58</td>
<td>Landi</td>
<td>204115_at</td>
</tr>
<tr>
<td>GNG11</td>
<td>–3.485</td>
<td>7.75E-11</td>
<td>1</td>
<td>19 : 20</td>
<td>Stearman</td>
<td>37908_at</td>
</tr>
<tr>
<td></td>
<td>–20.213</td>
<td>2.51E-11</td>
<td>1</td>
<td>17 : 132</td>
<td>Bhattacharjee</td>
<td>37908_at</td>
</tr>
<tr>
<td></td>
<td>–6.202</td>
<td>6.90E-25</td>
<td>2</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1782419</td>
</tr>
<tr>
<td></td>
<td>–2.830</td>
<td>4.15E-11</td>
<td>3</td>
<td>10 : 86</td>
<td>Beer</td>
<td>U31384_at</td>
</tr>
<tr>
<td></td>
<td>–4.296</td>
<td>3.94E-4</td>
<td>4</td>
<td>5 : 5</td>
<td>Wachi</td>
<td>204115_at</td>
</tr>
<tr>
<td></td>
<td>–2.181</td>
<td>7.18E-14</td>
<td>2</td>
<td>25 : 25</td>
<td>Wei</td>
<td>220032_at</td>
</tr>
<tr>
<td>CPED1</td>
<td>–3.108</td>
<td>1.13E-17</td>
<td>2</td>
<td>65 : 45</td>
<td>Hou</td>
<td>228728_at</td>
</tr>
<tr>
<td></td>
<td>–3.325</td>
<td>1.03E-34</td>
<td>1</td>
<td>58 : 58</td>
<td>Selamat</td>
<td>ILMN_1677038</td>
</tr>
</tbody>
</table>
Figure 8: Analysis of ZDHHC9, BTNL9, GNG11 and CPED1 in clinical lung adenocarcinoma patients using bioinformatics databases. The gene expression of GNG11 (A) and CPED1 (B), comparing 31 pairs of clinical lung adenocarcinoma (red) and adjacent normal tissue (blue), was performed on a GSE10072 microarray from the GEO database. The p-value of gene expression was calculated by t-test with Wilcoxon matched-pairs signed rank test. *** represents p < 0.001. The survival curves comparing 2 populations with high (red) and low (black) gene expression in lung adenocarcinoma patients were performed on the Kaplan–Meier plotter database – ZDHHC9 (C), BTNL9 (D) and CPED1 (I), and the PrognoScan database – BTNL9 (E, F), GNG11 (G, H), and CPED1 (J, K). The analysis criteria of the PrognoScan and Kaplan–Meier plotter databases were Cox p-value < 0.05 and log-rank p-value < 0.05 respectively. Raw data of the GEO and PrognoScan databases were extracted and re-plotted by GraphPad Prism 5 software. (L) The forest plots showed hazard ratios (95% CI, confidence interval) identified from the PrognoScan and Kaplan–Meier plotter databases.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Dataset</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZDHHC9</td>
<td>KM Plotter</td>
<td>0.67</td>
<td>[0.53-0.88]</td>
<td>230156_at</td>
</tr>
<tr>
<td></td>
<td>GSE31210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iacob-00182-MSK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MICHIGAN-LC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTNL9</td>
<td>KM Plotter</td>
<td>0.44</td>
<td>[0.29-0.66]</td>
<td>1553279_at</td>
</tr>
<tr>
<td></td>
<td>GSE31210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iacob-00182-MSK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MICHIGAN-LC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNG11</td>
<td>KM Plotter</td>
<td>0.67</td>
<td>[0.53-0.88]</td>
<td>230156_at</td>
</tr>
<tr>
<td></td>
<td>GSE31210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iacob-00182-MSK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MICHIGAN-LC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPED1</td>
<td>KM Plotter</td>
<td>0.67</td>
<td>[0.53-0.88]</td>
<td>230156_at</td>
</tr>
<tr>
<td></td>
<td>GSE31210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iacob-00182-MSK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MICHIGAN-LC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
GNG11 is a lipid-anchored protein, which has been reported to inhibit cell growth [71] and regulate cellular senescence in lymphoma [72]. We found that GNG11 may exert suppressor functions in the tumorigenesis of lung adenocarcinoma.

CPED1, also known as C7orf58, contains a cadherin-like beta sandwich domain [73], but the molecular function of CPED1 is unclear. Our data showed that CPED1 may play a role as a potential tumor suppressor in lung adenocarcinoma.

The summary of differentially expressed genes in the Oncomine database is shown in Supplementary Table 3. Seven of a total of 11 datasets showed similar patterns of genetic expression, suggesting that this molecular change is constant between lung adenocarcinoma and normal lung tissue. Thus, those genes found in this study may represent a novel gene expression signature in lung adenocarcinoma (Figure 10). The increased expression of AGR2 and decreased expression of IL33 have also been identified in other reports. We also analyzed the expression of microRNAs in lung adenocarcinoma (Table 9). Twenty-too upregulated microRNAs were identified in lung adenocarcinoma. We focused on microRNAs with predictable putative targets - BTNL9, FMO2, IL33, CPED1, and PDK4. Among these microRNAs, elevated expression of hsa-miR-183-5p [74], hsa-miR-33b-5p [75], hsa-miR-429 [76], hsa-miR-182-5p [77], and hsa-miR-130b-5p [78] have been associated with tumorigenesis in lung cancer. The function of hsa-miR-542-3p is unclear. However, since the genetic interactions of hsa-miR-183-5p-BTNL9, hsa-miR-33b-5p-CPED1, hsa-miR-429-CPED1, hsa-miR-182-5p-FMO2, hsa-miR-130b-5p-IL33, and hsa-miR-542-3p-IL33 have not been identified, these altered genetic regulations may play important roles in the progression of lung adenocarcinoma.

MATERIALS AND METHODS

Clinical lung adenocarcinoma specimens

Three pairs of tumors and adjacent non-tumor lung tissues were collected from the Division of Thoracic surgery and Division of Pulmonary and Critical Care Medicine, Kaohsiung Medical University Hospital (KMUH), Kaohsiung, Taiwan. Approval for these studies was obtained from the Institutional Review Board (IRB)

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>Precursor</th>
<th>RPM (Reads per million)</th>
<th>T/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-1307-5p</td>
<td>hsa-mir-1307</td>
<td>118.34</td>
<td>396.09</td>
</tr>
<tr>
<td>hsa-miR-130b-5p</td>
<td>hsa-mir-130b</td>
<td>2.31</td>
<td>12.19</td>
</tr>
<tr>
<td>hsa-miR-130b-3p</td>
<td>hsa-mir-130b</td>
<td>15.6</td>
<td>132.83</td>
</tr>
<tr>
<td>hsa-miR-182-5p</td>
<td>hsa-mir-182</td>
<td>926.7</td>
<td>5350.27</td>
</tr>
<tr>
<td>hsa-miR-183-5p</td>
<td>hsa-mir-183</td>
<td>179.58</td>
<td>1197.91</td>
</tr>
<tr>
<td>hsa-miR-190a-5p</td>
<td>hsa-mir-190a</td>
<td>6.7</td>
<td>30.48</td>
</tr>
<tr>
<td>hsa-miR-200a-5p</td>
<td>hsa-mir-200a</td>
<td>7.4</td>
<td>28.83</td>
</tr>
<tr>
<td>hsa-miR-200b-3p</td>
<td>hsa-mir-200b</td>
<td>459.71</td>
<td>1338.61</td>
</tr>
<tr>
<td>hsa-miR-21-3p</td>
<td>hsa-mir-21</td>
<td>472.88</td>
<td>2282.29</td>
</tr>
<tr>
<td>hsa-miR-224-5p</td>
<td>hsa-mir-224</td>
<td>15.95</td>
<td>85.34</td>
</tr>
<tr>
<td>hsa-miR-301b-3p</td>
<td>hsa-mir-301b</td>
<td>4.74</td>
<td>173.22</td>
</tr>
<tr>
<td>hsa-miR-31-5p</td>
<td>hsa-mir-31</td>
<td>91.18</td>
<td>452.09</td>
</tr>
<tr>
<td>hsa-miR-33b-5p</td>
<td>hsa-mir-33b</td>
<td>5.78</td>
<td>20.57</td>
</tr>
<tr>
<td>hsa-miR-345-5p</td>
<td>hsa-mir-345</td>
<td>34.44</td>
<td>97.53</td>
</tr>
<tr>
<td>hsa-miR-424-3p</td>
<td>hsa-mir-424</td>
<td>2.43</td>
<td>30.22</td>
</tr>
<tr>
<td>hsa-miR-424-5p</td>
<td>hsa-mir-424</td>
<td>35.82</td>
<td>240.9</td>
</tr>
<tr>
<td>hsa-miR-429</td>
<td>hsa-mir-429</td>
<td>143.41</td>
<td>407.01</td>
</tr>
<tr>
<td>hsa-miR-450a-5p</td>
<td>hsa-mir-450a-1</td>
<td>4.39</td>
<td>25.53</td>
</tr>
<tr>
<td>hsa-miR-450a-5p</td>
<td>hsa-mir-450a-2</td>
<td>4.39</td>
<td>25.4</td>
</tr>
<tr>
<td>hsa-miR-452-5p</td>
<td>hsa-mir-452</td>
<td>11.79</td>
<td>36.95</td>
</tr>
<tr>
<td>hsa-miR-542-3p</td>
<td>hsa-mir-542</td>
<td>11.33</td>
<td>54.23</td>
</tr>
<tr>
<td>hsa-miR-7705</td>
<td>hsa-mir-7705</td>
<td>3</td>
<td>14.6</td>
</tr>
<tr>
<td>hsa-miR-96-5p</td>
<td>hsa-mir-96</td>
<td>25.31</td>
<td>64</td>
</tr>
</tbody>
</table>
Figure 9: Identification of differentially expressed microRNAs in lung adenocarcinoma compared to adjacent normal tissue using next-generation sequencing. Venn diagram analysis showed 22 upregulated microRNAs (A) and 0 downregulated microRNAs (B) in lung adenocarcinoma, compared to adjacent normal tissue from 3 pairs of clinical specimens. The criteria were fold change > 2 (tumor/normal) and reads per million (RPM) > 1. (C) The heatmap diagram showed the differentially expressed genes with z-score (log2) values by using color clustering on the GENE-E web-tool. Green represents downregulation (minimum = –3.0), and red represents upregulation (maximum = 3.0). (D) The “Targets” Venn diagram shows the predicted genes of microRNAs from the “microRNAs” Venn diagram using the miRmap web-site database. The selection threshold was miRmap score ≥ 90.0. The intersection Venn diagram between “mRNAs” and “Targets” showed total of 6 potential microRNA-mRNA interactions.
of KMUH, and informed consent was obtained from all patients in accordance with the Declaration of Helsinki.

Next-generation sequencing (NGS)

The expression profile of mRNA and microRNA was performed using NGS [22]. Three pairs of lung adenocarcinomas and adjacent normal specimens were used in this project. Total RNA was extracted by using Trizol® Reagent (Invitrogen, USA), according to the manufacturer’s instructions. The cell lysates were applied to Welgene Biotechnology Company (Welgene, Taipei, Taiwan) for RNA-seq and small RNA-seq analysis. The criteria for differentially expressed mRNA analysis were fold change > 2 and fragments per kilobase million (FPKM) > 0.3. The criteria for differentially expressed microRNAs' selection were fold change > 2 and reads per million (RPM) > 1.

Oncomine database analysis

The Oncomine database contains over 18,000 microarray experiments and 35 major cancer types [23]. The raw data of mRNA expression in clinical lung adenocarcinoma and normal specimens (cancer vs. normal) were extracted from the Oncomine database (http://www.oncomine.org, Compendia biosciences, Ann Arbor, MI, USA). The criteria in the analysis were $P$-value < 1E-4, fold change > 2, and gene rank in top 10%. $P$-value was calculated using the Oncomine database through two-sided Student's $t$-test. For the comparison of genes in each dataset, raw data was extracted and re-plotted using the GENE-E web-tool (https://software.broadinstitute.org/GENE-E/), and the relative color scheme was used for clustering as minimum = 0 (blue) and maximum = 1 (yellow). Eleven datasets were selected.

Table 10: Genes selected between differentially expressed genes and putative targets of microRNA

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Predicted targets</th>
<th>miRmap score</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-183-5p</td>
<td>BTNL9</td>
<td>92.2457</td>
</tr>
<tr>
<td>hsa-miR-200b-3p</td>
<td>CPED1</td>
<td>97.3127</td>
</tr>
<tr>
<td>hsa-miR-33b-5p</td>
<td>CPED1</td>
<td>91.5414</td>
</tr>
<tr>
<td>hsa-miR-429</td>
<td>CPED1</td>
<td>97.7498</td>
</tr>
<tr>
<td>hsa-miR-182-5p</td>
<td>FMO2</td>
<td>98.0884</td>
</tr>
<tr>
<td>hsa-miR-345-5p</td>
<td>FMO2</td>
<td>96.6797</td>
</tr>
<tr>
<td>hsa-miR-130b-5p</td>
<td>IL33</td>
<td>98.1906</td>
</tr>
<tr>
<td>hsa-miR-542-3p</td>
<td>IL33</td>
<td>96.593</td>
</tr>
<tr>
<td>hsa-miR-21-3p</td>
<td>NDRG4</td>
<td>95.067</td>
</tr>
<tr>
<td>hsa-miR-424-5p</td>
<td>PDK4</td>
<td>99.1986</td>
</tr>
</tbody>
</table>

Figure 10: The proposed novel molecular signatures of gene regulations involved in lung adenocarcinoma.
for our analysis, including Hou (normal = 65 and lung adenocarcinoma = 45) [79], Landi (normal = 49 and lung adenocarcinoma = 58) [80], Selamat (normal = 58 and lung adenocarcinoma = 58) [81], Okayama (normal = 20 and lung adenocarcinoma = 226) [82], Su (normal = 30 and lung adenocarcinoma = 27) [83], Wei (normal = 25 and lung adenocarcinoma = 25) [84], Stearman (normal = 19 and lung adenocarcinoma = 20) [85], Bhattacharjee (normal = 17 and lung adenocarcinoma = 132) [86], Beer (normal = 10 and lung adenocarcinoma = 86) [87], Garber (normal = 5 and lung adenocarcinoma = 40) [88], and Wachi (normal = 5 and lung adenocarcinoma = 5) [89].

SurvExpress database analysis

SurvExpress integrates the TCGA database (https://tcga-data.nci.nih.gov/) which provides microarray information, including cancer type, survival, and gene expression. The correlation between IL33 mRNA expression and survival rate was analyzed on the SurvExpress web-database (http://bioinformatica.itysm.mx/SurvExpress). The dataset lung adenocarcinoma TCGA (N = 255) was used in our analysis. Samples of each dataset were split into 2 risk groups (high and low risk) of the same size, of which each group was determined by the ordered Prognostic Index (PI, high value for high risk) [27]. Prognostic Index (PI) is the linear component of the Cox model, computed by gene expression value multiplied with values estimated from the Cox fitting [90].

PrognoScan database analysis

PrognoScan collects information of the GEO (Gene Expression Omnibus) microarray database, including cancer type, survival rates, and gene expression. The correlation between gene expression and overall survival rates was performed on the PrognoScan web-database (http://www.abren.net/PrognoScan/) [25]. The raw data were extracted and re-plotted by GraphPad Prism 5 software (GraphPad Software, La Jolla, CA, USA). The threshold was determined as Cox p-value < 0.05. Samples of each dataset were divided into 2 expression groups (high and low) at the potential cutpoint. The cutpoint (from < 0.1 or > 0.9 quantile) was estimated by the minimum P-value approach [91], and the P-value correlation was calculated by the formula [92]. The hazard ratios (95% confidence intervals) of each dataset was calculated using the Cox proportional model, and are listed in the related Tables. HR = 0 represents no difference between 2 groups, HR < 1 represents better survival rate in the population with higher levels of expression, and HR > 1 represents better survival rates in the population with lower levels of expression. The specific probe of each dataset is listed in its related Figure.

Kaplan–Meier plotter database analysis

The Kaplan–Meier plotter is a web-database providing the information on 54675 genes’ expression and survival rates in 10461 cancer samples, including 5143 breast, 1816 ovarian, 2437 lung, and 1065 gastric cancer patients [26]. The correlation of gene expression and overall survival rates in lung cancer was determined, and lung adenocarcinoma (N = 720) was selected in our analysis. Patients were split into 2 populations with the best cut-off, which was computed with median survival. The hazard ratios (95% confidence intervals) were calculated using the Cox proportional model, and are listed in the related Tables. The specific probe of each dataset was listed in its related Figure.

Gene expression omnibus (GEO) database analysis

GEO is a web-database providing submitted high throughput gene expression data of microarrays, chips, or NGS (https://www.ncbi.nlm.nih.gov/geo/) [24]. We selected the microarray with accession number GSE10072, published in 2008 [80], for this project. This microarray provides gene expression information of 180 clinical lung adenocarcinoma and non-tumor samples. Here, we selected 31 pairs of lung adenocarcinoma with adjacent normal tissue for gene expression analysis. The raw data were analyzed and extracted from GEO2R (https://www.ncbi.nlm.nih.gov/geo2r/), and re-plotted by GraphPad prism 5 software (GraphPad Software, La Jolla, CA, USA). The p-value was calculated by using t-test with Wilcoxon matched-pairs signed rank test.

miRmap database analysis

miRmap is a web-tool database providing analysis of microRNA targets prediction (http://mirmap.ezlab.org/) [28]. It identifies the putative target genes by calculating the complementary ability of microRNA-mRNA interactions. The strength of mRNA repression is estimated for ranking potential candidate targets by employing various features, including thermodynamic, evolutionary, probabilistic or sequence-based features. The prediction results show a list of putative target genes with miRmap scores, which are a predictive reference values. Putative targets with miRmap scores ≥ 90.0 were selected for this project.

Statistical analysis

The raw data extracted from GEO database were statistically analyzed using t-test with Wilcoxon matched-pairs signed rank test by GraphPad Prism 5 software (GraphPad Software, La Jolla, CA, USA).
ACKNOWLEDGMENTS AND FUNDING

This study was supported by grants from the Ministry of Science and Technology (MOST 106-2314-B-037-064; MOST 104-2320-B-037-014-MY3; MOST 104-2314-B-037-053-MY4; MOST 103-2320-B-037-006-MY3), the Kaohsiung Medical University “Aim for the Top 500 Universities Grant” (Grant No. KMU-TP105C05), the Kaohsiung Medical University “Aim for the Top Journals Grant” (Grant No. KMU-DT106005), and the Kaohsiung Medical University Hospital Research Foundation (KMUH106-6R13).

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES


